



⑮ **BUNDESREPUBLIK  
DEUTSCHLAND**



**DEUTSCHES  
PATENT- UND  
MARKENAMT**

⑫ **Offenlegungsschrift**  
⑩ **DE 100 38 694 A 1**

⑤ Int. Cl.<sup>7</sup>:  
**C 12 Q 1/04**  
G 01 N 27/64

⑳ Aktenzeichen: 100 38 694.6  
㉔ Anmeldetag: 28. 7. 2000  
㉕ Offenlegungstag: 14. 2. 2002

**DE 100 38 694 A 1**

㉑ **Anmelder:**  
AnagnosTec Gesellschaft für analytische  
Biochemie und Diagnostik mbH, 14943  
Luckenwalde, DE  
  
㉒ **Vertreter:**  
Schneider, H., Dipl.-Ing., Pat.-Anw., 10117 Berlin

㉓ **Erfinder:**  
Kallow, Wibke, Dr., 12279 Berlin, DE; Dieckmann,  
Ralf, Dr., 10707 Berlin, DE; Kleinkauf, Niels, Dr.,  
04229 Leipzig, DE; Erhard, Marcel, Dr., 14109 Berlin,  
DE; Neuhof, Torsten, 14197 Berlin, DE  
  
㉔ **Entgegenhaltungen:**  
WO 00 29 987 A1

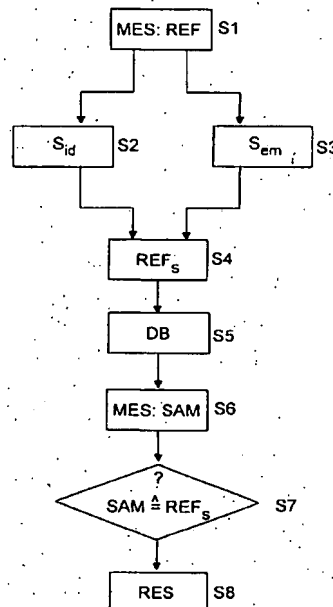
**Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen**

Prüfungsantrag gem. § 44 PatG ist gestellt

㉕ **Verfahren zur Identifizierung von Mikroorganismen mittels MALDI-TOF-MS**

㉖ Die Erfindung betrifft ein Verfahren zur Identifizierung von Mikroorganismen mittels Matrix-assisted-Laser-Desorption-Ionisation Time-of-Flight Massenspektrometrie (MALDI-TOF-MS), wobei eine Datenbank mit Referenzspektren, bestehend aus Massenspektren einer Anzahl bekannter Mikroorganismen, erstellt wird, ein MALDI-TOF-Massenspektrum (Probenspektrum) einer Probe eines zu identifizierenden Mikroorganismus aufgenommen wird und eine Ähnlichkeitsanalyse des Massenspektrums des zu identifizierenden Mikroorganismus mit den in der Datenbank enthaltenen Referenzspektren durchgeführt wird.

Es ist vorgesehen, dass als Referenzspektren (REF) synthetische Massenspektren (REF<sub>S</sub>) verwendet werden, die durch Zusammenfassung einer gegenüber natürlichen Massenspektren reduzierten Anzahl von für den jeweiligen Mikroorganismus spezifischen Signalen (S) erzeugt werden.



**DE 100 38 694 A 1**

[0001] Die Erfindung betrifft ein Verfahren zur Identifizierung von Mikroorganismen mittels Matrix-assisted-Laser-Desorption-Ionisation-Time-of-Flight Massenspektrometrie (MALDI-TOF-MS) mit den im Oberbegriff des Anspruchs 1 genannten Merkmalen sowie eine nach dem Verfahren erstellte und für das Verfahren verwendbare Datenbank.

[0002] Eine schnelle und zuverlässige Identifizierung von Mikroorganismen ist in verschiedenen Bereichen des Gesundheitswesens, beispielsweise der Diagnostik von Infektionen, sowie der Lebensmittelindustrie von entscheidender Bedeutung. Die traditionelle Identifizierung mittels des direkten Bakteriennachweises erfordert zunächst die Anzucht der zu identifizierenden Mikroorganismen aus einer Materialprobe. Anschließend mikroskopische Untersuchungsverfahren dienen hauptsächlich der vorläufigen Orientierung über den Bakteriengehalt sowie der Mikromorphologie beziehungsweise Färbeeigenschaften der klinischen Untersuchungsprobe. Die Identifizierung von Isolaten bis zur Speziesebene erfordert nicht selten eine Subkultivierung zwecks der Gewinnung einer Reinkultur. Nach dem klassischen Identifizierungsverfahren der medizinischen Mikrobiologie werden schließlich unter Verwendung einer geeigneten Kombination von Differentialmedien (sogenannte "Bunte Reihe"), spezifische Stoffwechselleistungen des zu identifizierenden Mikroorganismus erfasst. Hauptnachteil des mikrobiellen Verfahrens ist sein sehr hoher Zeitbedarf. Modernere molekularbiologische Ansätze, wie die PCR-Methode (Polymerase-Kettenreaktion) und die 16S-rRNA-Methode, involvieren die genetische Analyse des zuvor isolierten Genoms beziehungsweise bestimmter Ribonukleinsäuren. Diese Verfahren haben wegen ihrer hohen Empfindlichkeit stark an Bedeutung gewonnen. Sie erfordern ebenso wie die mikrobiologische Charakterisierung eine Kultivierung der Organismen im Labor und sind darüber hinaus mit einem erheblichen personellen und instrumentellen Aufwand belastet. Ferner ist ein infrarotspektroskopisches Verfahren entwickelt worden, bei dem Schwingungsspektren intakter Zellen ("Fingerprintspektren") in FTIR-Spektrometern aufgezeichnet und mit einer Datenbank mit Schwingungsspektren bekannter Mikroorganismen abgeglichen wird. Diese noch sehr neue Technik befindet sich noch in der Entwicklung und kann derzeit nur von speziell geschultem und erfahrenem Personal durchgeführt werden, so dass sich dieser Ansatz noch nicht in der Praxis durchsetzen konnte.

[0003] Mit der sogenannten MALDI-TOF-Massenspektrometrie ist in den letzten Jahren ein Verfahren entwickelt worden, das im Gegensatz zu herkömmlichen massenspektrometrischen Verfahren auch der Analyse biologischer Makromoleküle zugänglich ist. Bei der MALDI-TOF-MS-Technik wird die zu untersuchende Probe mit einer meist kristallinen organischen Verbindung, der sogenannten Matrix, auf eine Probenplatte gegeben, wobei die Probe in die Matrixkristalle eingebaut wird, und mit einem Laserstrahl in Wechselwirkung gebracht. Dabei werden einzelne Moleküle der Probe von dem Probeträger desorbiert und ionisiert. Anschließend werden die so erzeugten Ionen in einem elektrischen Feld beschleunigt und ihre Flugzeit (Time-of-Flight) bis zum Erreichen eines Detektors registriert. Da die Beschleunigung von der Masse eines ionisierten Moleküls abhängt, spiegeln die Flugzeiten die in der Probe vorhandenen Molekülmassen wider. Die MALDI-TOF-MS-Technik findet heutzutage hauptsächlich auf dem Gebiet der Proteinanalytik ("Proteomics") und bei der RNA- und DNA-Analytik Anwendung. Es wurde vorgeschlagen, die Methode zur Identifizierung von Mikroorganismen in Analogie zu der

oben beschriebenen infrarotspektroskopischen Fingerprintmethode einzusetzen. Dafür wird das MALDI-TOF-Massenspektrum eines Zellextraktes oder intakter Zellen (zum Beispiel WO 98/09314) des unbekannten Mikroorganismus mit den Spektren bekannter Organismen verglichen. Der Vergleich des Probenspektrums mit den Referenzspektren der Datenbank erfolgt in der Regel computergestützt mittels statistisch-mathematischer Algorithmen, die in sogenannten Mustererkennungsverfahren entwickelt wurden. Die Massenspektren ähneln sich mit wachsender verwandtschaftlicher Nähe der Mikroorganismen zunehmend. Das heißt, die Wahrscheinlichkeit einer Wiederholung bestimmter Signale in den Massenspektren nimmt zu. Daher ist das Verfahren bei übergeordneten Klassifizierungsniveaus, etwa bei Gattungen oder Familien, mit einer erhöhten Unsicherheit verbunden, wenn eine Zuordnung auf Stammniveau fehlschlägt, beispielsweise weil kein Referenzspektrum des Stammes in der Datenbank vorhanden ist. Das Verfahren ist daher auf sehr umfangreiche Datenbanken mit einer Vielzahl repräsentativer Referenzspektren von bekannten Stämmen angewiesen. Hinzu kommt die Schwierigkeit, ein spektrales Grundrauschen von echten Signalen zu diskriminieren.

[0004] Aufgabe der vorliegenden Erfindung ist es daher, das gattungsgemäße MALDI-TOF-MS-Verfahren zur Identifizierung von Mikroorganismen dahingehend weiterzuentwickeln, dass seine Zuverlässigkeit erhöht wird, ohne den erforderlichen Prozessaufwand zu vergrößern. Es soll ferner eine nach dem Verfahren erstellte und für das Verfahren einsetzbare Datenbank zur Verfügung gestellt werden.

[0005] Diese Aufgabe wird durch ein Verfahren mit den Merkmalen des Anspruchs 1 sowie durch eine Datenbank und ihre Verwendung nach den Ansprüchen 13 und 14 gelöst.

[0006] Verfahrensgemäß ist vorgesehen, dass als Referenzspektren synthetische Massenspektren verwendet werden, die durch Zusammenfassung einer gegenüber "natürlichen" (nicht reduzierten) Massenspektren reduzierten Anzahl von für den jeweiligen Mikroorganismus spezifischen Signalen erzeugt werden. Durch die Reduzierung der Referenzspektren auf eine verhältnismäßig geringe Anzahl charakteristischer Signale, wird eine erhebliche Daten- und Informationsreduktion erreicht. Durch die Datenreduktion wird nicht nur Speicherplatz eingespart, sondern auch ein zeitlicher Aufwand für Datenübertragungen, so dass das Verfahren prinzipiell auch für eine Anwendung über vernetzte Datenverarbeitungsanlagen (z. B. Internet) geeignet ist. Darüber hinaus ermöglicht die Informationsreduktion der Referenzspektren der Datenbank eine deutliche Beschleunigung der Ähnlichkeitsanalyse des Massenspektrums des zu identifizierenden Mikroorganismus mit den Referenzspektren, da die Analyse sich nunmehr auf einen Vergleich der in den Referenzspektren enthaltenen Signale beschränken kann.

[0007] Verglichen mit herkömmlichen Verfahren, in denen ein Probenspektrum des zu identifizierenden Organismus mit "natürlichen" Referenzspektren abgeglichen wird, ist eine Zuverlässigkeit des erfindungsgemäßen Verfahrens, das heißt die Wahrscheinlichkeit, eine korrekte Zuordnung des unbekannten Organismus zu treffen, deutlich erhöht. Hierin ist der bedeutendste Vorteil der Erfindung zu sehen. Die erhöhte Zuverlässigkeit ist durch die hohe Konzentrierung spezifischer Informationen in den Referenzspektren zu erklären, die nicht durch Signale geringer Signifikanz oder Rauschen überdeckt wird. Auch bei einem möglichen Versagen bei einer Zuordnung auf Stammebene einer Probe, etwa weil kein Referenzspektrum dieses Stammes in der Datenbank verfügbar ist, liefert das Verfahren zuverlässige Identi-

fizierungen auf übergeordneten Klassifizierungsebenen, beispielsweise auf Gattungs- oder Artenebene. Die Empfindlichkeit der erfindungsgemäßen Vorgehensweise wird auch nicht durch Unterschiede in verschiedenen Spektren beeinträchtigt, welche beispielsweise durch unterschiedliche Kultivierungsbedingungen, unterschiedliche Zellstadien oder abweichende Signal-Rauschverhältnisse verursacht werden. [0008] Ein weiterer Vorteil des Verfahrens stellt die geringe erforderliche Probenmenge des zu identifizierenden Mikroorganismus dar, die in kürzeren Kultivierungszeiten erhältlich ist. Ferner können auch Mischkulturen analysiert werden, so dass auf die Anzucht von Reinkulturen verzichtet werden kann.

[0009] Die in den synthetischen Referenzspektren enthaltenen Signale können in zwei Kategorien unterschieden werden. Eine besonders vorteilhafte Ausgestaltung der Erfindung sieht vor, dass die Signale eines Referenzspektrums mindestens ein identifiziertes Signal umfassen, welches einem charakterisierten molekularen Zellbestandteil des jeweiligen Mikroorganismus eindeutig zugeordnet wurde. Zur Identifizierung eines Mikroorganismus geeignete Zellbestandteile sind beispielsweise bestimmte Peptide, Proteine, Ribonukleinsäuren und/oder Lipide. Es hat sich als besonders vorteilhaft erwiesen, eine Signalzuordnung eines ribosomalen Proteins, insbesondere eines Proteins der großen Ribosomenuntereinheit, vorzunehmen. Für Bakterien sind dies Proteine der sogenannten 50S-Untereinheit und bei Pilzen die 60S-Untereinheit. Proteine stellen einen Hauptbestandteil mikrobieller Zellen dar. Dies gilt insbesondere für Ribosomenproteine, die unabhängig von einem Entwicklungsstadium der Zelle, einem Nährstoffangebot oder sonstiger Kultivierungsbedingungen ständig präsent sind und somit zuverlässige Signale in den Massenspektren darstellen. Hinzu kommt, dass Aminosäuresequenzen analoger Proteine unterschiedlicher Spezies oder sogar unterschiedlicher Stämme sich zumindest geringfügig voneinander unterscheiden. Folglich weisen die analogen Proteine unterschiedliche Massen auf und eignen sich für deren Unterscheidung. So gelang den Erfindern beispielsweise erstmalig die Zuordnung dreier Massensignale in dem Massenspektrum von *Escherichia coli* ( $m/z = 4365, 6413, 7276$ ) zu den Proteinen L29, L30 und L36 der großen 50S-Ribosomenuntereinheit. Diese Signale sind ständige Bestandteile in Massenspektren von *E. coli*, nicht jedoch von den meisten anderen Mikroorganismen. Sie sind daher besonders zur Identifizierung von *Escherichia coli* und zur Aufnahme in ein synthetisches Referenzspektrum für *E. coli* geeignet.

[0010] In einer weiteren Ausbildung der vorliegenden Erfindung ist vorgesehen, dass die Signale eines synthetischen Referenzspektrums als zweite Signalkategorie mindestens ein empirisches Signal umfassen, welches durch Vergleich einer Vielzahl von Massenspektren bekannter Mikroorganismen als spezifisch für einen Mikroorganismus ermittelt wurde. Dabei handelt es sich um Signale, deren Ursprung, das heißt deren verursachende molekulare Zellkomponente, zwar nicht bekannt ist, die jedoch aufgrund bestimmter Kriterien als charakteristisch für einen Mikroorganismus gewertet werden. Vorzugsweise umfassen diese Kriterien eine vorgebbare Mindest-Häufigkeit für das Auftreten eines Signals in einer Anzahl von Massenspektren des selben Mikroorganismus sowie eine vorgebbare durchschnittliche Mindest-Intensität des Signals. Dabei sollte die Mindest-Häufigkeit mindestens 50%, insbesondere mindestens 70%, vorzugsweise mindestens 90%, betragen. Die Ermittlung der empirischen Signale durch Vergleich gemessener Massenspektren kann visuell, vorzugsweise jedoch computergestützt, durchgeführt werden. Entsprechende Algorithmen (Mustererkennungsverfahren) sind bekannt und sollen hier

nicht näher erläutert werden. Die Anzahl der Signale eines Referenzspektrums beträgt vorteilhafterweise 5 bis 30. In praktischen Versuchen hat sich insbesondere eine Anzahl von 10 bis 15 als ausreichend erwiesen. Es ist ferner bevorzugt vorgesehen, dass ein Signal eines synthetischen Referenzspektrums durch lediglich ein Koordinatenpaar repräsentiert wird. Dabei besteht das Koordinatenpaar aus einer Masse beziehungsweise einem Masse-Ladungs-Verhältnis als x-Koordinate einerseits und einer absoluten oder relativen Intensität als y-Koordinate andererseits. Verglichen mit "natürlichen" Massenspektren, die etwa 16000 Datenpunkte enthalten, bedeutet dies eine erhebliche Datenreduktion. Für den Abgleich eines Probespektrums mit den Referenzspektren der Datenbank können vorteilhafterweise den einzelnen identifizierten und empirischen Signalen der synthetischen Referenzspektren Gewichtungen entsprechend ihrer Signifikanz zugeordnet werden.

[0011] Die Erfindung umfasst ferner eine Datenbank, die mit dem erfindungsgemäßen Verfahren erstellt wurde, sowie die Verwendung der Datenbank für die Identifizierung von Mikroorganismen mittels MALDI-TOF-MS.

[0012] Die Erfindung wird nachfolgend in Ausführungsbeispielen anhand der zugehörigen Zeichnungen näher erläutert. Es zeigen:

[0013] Fig. 1 ein Ablaufschema einer typischen Ausführung des erfindungsgemäßen Verfahrens;

[0014] Fig. 2 MALDI-TOF-Massenspektren von *Escherichia coli* und von isolierten Ribosomen aus *Escherichia coli* und

[0015] Fig. 3 bis 6 synthetische Referenzspektren von zwei *Escherichia coli*-Stämmen sowie zweier *Klebsiella*-Spezies.

[0016] Fig. 1 zeigt in einem Fließschema einen typischen erfindungsgemäßen Verfahrensablauf. In einem ersten Schritt S1 wird eine möglichst große Anzahl von Referenzspektren REF bekannter Mikroorganismen gemessen. Im Folgenden werden einige Details zur Probenvorbereitung und zur Datenakquisition gegeben.

#### Probenvorbereitung und MALDI-TOF-Datenakquisition

[0017] Für eine MALDI-TOF-Analyse werden zirka 5 bis 100 µg Nasszellen oder auch 5 bis 50 µg getrocknete Zellen eines bekannten oder zu identifizierenden Bakteriums oder Pilzes benötigt. Eine Vorbehandlung der Zellen, beispielsweise ein Zellaufschluss, ist nicht notwendig. Die Nasszellen können entweder direkt von einer AGAR-Kultur mit einer sterilen Impföse auf eine auch Template genannte Probenplatte übertragen werden. Alternativ können auch abzentrifugierte Zellen einer Flüssigkultur verwendet werden. Anschließend werden die Zellen auf der Probenplatte mit 0,2 bis 1 µl einer Matrixlösung vermischt. In Abwandlung von dieser Prozedur können die Nasszellen auch vor ihrer Übertragung auf die Probenplatte mit der Matrixlösung vermischt und als Suspension übertragen werden. Für die folgenden Messungen wurde eine Matrixlösung aus 100 mg/ml 2,5-Dihydroxybenzoesäure in einer Mischung aus 50% Acetonitril und 50% Wasser mit 3% Trifluoressigsäure verwendet. Andere bekannte Matrixlösungen sind ebenfalls geeignet. Nach Trocknung der Probe, die mit einer Kristallbildung einhergeht, kann die Probe direkt einer MALDI-TOF-massenspektrometrischen Analyse unterworfen werden. Für die vorliegenden Messungen wurde jeder Probenpunkt einer Probenplatte in einem herkömmlichen Massenspektrometer mit zirka 50 bis 300 Laserpulsen eines Stickstoff-Lasers mit einer Wellenlänge von 337 nm angeregt. Die Akquisition positiver Ionenmassenspektren erfolgte im Linearmodus in einem Massenbereich von 2000 bis 20000  $m/z$ . Ein

typisches Beispiel eines auf diese Weise erhaltenen positiven MALDI-TOF-Massenspektrums REF von *Escherichia coli* ist im oberen Teil der Fig. 2 im Massenbereich von zirka  $m/z = 4000$  bis 14000 dargestellt.

[0018] Im anschließenden Schritt S2 (Fig. 1) wird eine Peakzuordnung zu bestimmten molekularen Zellbestandteilen durchgeführt. Eine mögliche Vorgehensweise, die auf bekannte Methoden der Biochemie und Molekularbiologie zurückgreift, wird im Folgenden am Beispiel von Ribosomenproteinen der großen Ribosomenuntereinheit kurz erläutert.

#### Identifizierung unbekannter Signale

[0019] Ein Proteinextrakt einer Zellkultur wird mittels einer 2D-Gelelektrophorese aufgetrennt. In Frage kommende sogenannte Proteinspots werden anschließend einem tryptischen Verdau, bei dem das Protein enzymatisch in kleine Proteinbruchstücke gespalten wird, unterzogen. Sofern Antikörper gegen Ribosomen vorhanden sind, können die relevanten Proteinspots auch durch ein Immunoassay (zum Beispiel Western-Blot-Analyse) erkannt werden. Die durch den tryptischen Verdau gewonnenen Proteinbruchstücke werden dann mittels einer HPLC, die für besonders kleine Probenmengen geeignet ist, aufgetrennt und ansequenziert. Mit den auf diese Weise ermittelten Sequenzbruchstücken kann dann versucht werden, die Ribosomengene in einer Datenbank zu identifizieren. Sind entsprechende Gene in der Datenbank vorhanden, kann aus der gesamten Gensequenz die korrespondierende Proteinmasse ermittelt werden. Kann ein entsprechendes Gen in der Datenbank nicht gefunden werden, so muss mit bekannten Mitteln der Molekularbiologie, die hier nicht näher ausgeführt werden sollen, das gesamte Gen, welches für das entsprechende ribosomale Protein kodiert, isoliert und sequenziert werden. Ist die Gensequenz bekannt, folgt die Übersetzung in die Proteinsequenz und Ermittlung der theoretischen Proteinmasse. Eine Überprüfung dieser theoretischen Masse kann erfolgen, indem der entsprechende Proteinspot der 2D-Gelelektrophorese einer MALDI-TOF-Massenspektrometrie unterworfen wird. Bei Abweichungen von der theoretischen Proteinmasse können Modifikationen des Proteins durch MALDI-TOF-Analyse des tryptischen Verdau festgestellt werden.

[0020] Zur Verdeutlichung ist in der Fig. 2 im unteren Teil ein Massenspektrum des 70S-Ribosoms aus *Escherichia coli*, das mittels 2D-Gelelektrophorese isoliert wurde, dargestellt. Mit 70S wird der gesamte Ribosomen eines Proteins bezeichnet, der sich aus der großen 50S-Untereinheit und der kleinen 30S-Untereinheit zusammensetzt. Beide Untereinheiten bestehen ihrerseits aus einer Reihe von Proteinen, die mit dem Buchstaben L (für large) und dem Buchstaben S (für small) bezeichnet werden. Es ist ohne Weiteres ersichtlich, dass eine Reihe der intensivsten Signale aus dem Spektrum von *Escherichia coli* (vgl. Fig. 2 oberer Teil) sich auf ribosomale Proteine zurückführen lassen. Entsprechende Signale sind in der Abbildung mit Pfeilen gekennzeichnet. Den Erfindern ist es erstmalig gelungen, drei dieser Signale bestimmten Proteinen der großen 50S-Ribosomenuntereinheit zuzuordnen. Dabei handelt es sich um die Proteine L29, L30 und L36 mit den Massen  $m/z = 7274$ , 6412 und 4364. Da diese Signale mit einer hohen Zuverlässigkeit in den Massenspektren von *Escherichia coli* auftauchen, sind sie besonders zur Identifizierung geeignet. Sie wurden als identifizierte Signale  $S_{id}$  für das synthetische Referenzspektrum  $REF_S$  für *Escherichia coli* verwendet.

[0021] In einem alternativen oder zusätzlichen Schritt S3 (Fig. 1) werden empirische Signale  $S_{em}$  aus den gemessenen Referenzspektren REF der bekannten Mikroorganismen ermittelt. Die Ermittlung der empirischen Signale  $S_{em}$  erfolgt

durch Vergleich einer Vielzahl von Massenspektren, die von demselben Stamm aufgezeichnet wurden, untereinander sowie durch Vergleich dieser Massenspektren mit denen von anderen Organismen. Dabei werden solche Signale als charakteristisch für einen Mikroorganismus ermittelt, die möglichst häufig in den Massenspektren desselben Organismus auftauchen und möglichst selten in den Massenspektren eines anderen. Für die Häufigkeit eines Auftretens eines geeigneten empirischen Signals  $S_{em}$  kann dabei ein Mindestwert, beispielsweise  $> 70\%$  bezogen auf alle Spektren eines Organismus, vorgegeben werden. Ebenso sollte ein solches Signal eine vorgebbare Mindest-Intensität besitzen, um seine Unterscheidung von dem Untergrundrauschen zu erleichtern. Die Ermittlung der empirischen Signale  $S_{em}$  im Schritt S3 kann durch visuellen Spektrenvergleich durch den Anwender erfolgen. Es ist jedoch bevorzugt vorgesehen, diesen Schritt automatisiert mit Hilfe geeigneter Rechnerprogramme durchzuführen. Hier kommt etwa ein Algorithmus in Frage, welcher die gemessenen Referenzspektren REF hinsichtlich der genannten Kriterien untersucht. Es sind jedoch auch, beispielsweise aus der Infrarotspektrometrie, computergestützte Verfahren bekannt, die unter Anwendung statistischer Algorithmen in der Lage sind, wiederkehrende Signale zu erkennen und herauszufiltern.

[0022] Die in den Schritten S2 und S3 ermittelten identifizierten und empirischen Signale  $S_{id}$ ,  $S_{em}$  werden in einem anschließenden Schritt S4 zu synthetischen Referenzspektren  $REF_S$  zusammengefasst. Dabei wird für jeden bekannten Mikroorganismus ein Referenzspektrum  $REF_S$  erzeugt. In den Fig. 3 bis 6 sind beispielhaft derartige synthetische Referenzspektren  $REF_S$  für zwei *Escherichia coli*-Stämme (ATCC 25922 und ATCC 35218) und zwei *Klebsiella*-Spezies – nämlich *Klebsiella pneumoniae* und *Klebsiella oxytoca* – dargestellt. Dabei ist jeweils im oberen Teil jeder der Fig. 3 bis 6 das synthetische Referenzspektrum  $REF_S$  in Koordinatenform dargestellt, während im unteren Teil jeweils die graphische Repräsentation in der typischen Form von Massenspektren abgebildet ist. Die synthetischen Referenzspektren  $REF_S$  umfassen in den gezeigten Beispielen elf bis vierzehn Signale  $S$ . Dabei sind in den *Escherichia coli*-Referenzspektren  $REF_S$  die als die Proteine L29, L30 und L36 der großen 50S-Ribosomenuntereinheit von *Escherichia coli* identifizierten Signale  $S_{id}$  markiert. Diese Signale  $S_{id}$  tauchen in beiden Stämmen von *Escherichia coli* auf. Sie eignen sich zur Identifizierung der Art und zur Abgrenzung gegenüber *Klebsiella*-Spezies, in denen nur ein Teil der identifizierten Signale  $S_{id}$  vorkommt. Insbesondere wird bei *Klebsiella* (Fig. 5 und 6) das Signal mit der Masse  $m/z = 6413$  nicht beobachtet. Die Unterscheidung der beiden *Escherichia coli*-Stämme (Fig. 3 und 4) wird durch solche empirische Signale  $S_{em}$  ermöglicht, die nur bei einem von beiden beobachtet werden. Ein Vergleich der in den Fig. 3 und 4 gezeigten synthetischen Referenzspektren  $REF_S$  zeigt, dass das erfindungsgemäße Verfahren empfindlich genug ist, um eine Unterscheidung von Mikroorganismen sogar auf Stammniveau zu ermöglichen. Daneben wird auch eine Unterscheidung zwischen *Escherichia coli* einerseits und *Klebsiella*-Spezies andererseits ermöglicht.

[0023] Die in dem Schritt S4 (Fig. 1) erzeugten synthetischen Referenzspektren  $REF_S$  werden in einem anschließenden Schritt S5 zu einer Datenbank DB zusammengefasst. Innerhalb der Datenbank DB können die synthetischen Referenzspektren  $REF_S$  in logischer Weise, beispielsweise nach Familie, Gattung, Art und Stamm, geordnet werden. Der Verfahrensablauf von S1 bis S5, das heißt die Erstellung der Datenbank DB, muss prinzipiell nur einmal durchgeführt werden. Auf diese Datenbank DB kann dann bei den anschließenden Schritten immer wieder zurückge-

griffen werden. Selbstverständlich sollte die Datenbank DB ständig durch Referenzspektren  $REF_s$  weiterer Mikroorganismen erweitert und aktualisiert werden. So ist etwa denkbar, Referenzspektren  $REF_s$  von Mutanten, insbesondere solchen mit verändertem Resistenz- und/oder Virulenzverhalten, aufzunehmen, sofern diese Stämme charakteristische Signale liefern. Auch sollten auf Basis statistischer Auswertung der analysierten Proben die vorhandenen Referenzspektren  $REF_s$  laufend optimiert werden, beispielsweise durch die Aufnahme neuer Signale  $S_{id}$ ,  $S_{em}$  und/oder durch Löschung von Signalen.

[0024] In dem Schritt S6 erfolgt dann die Akquisition eines Massenspektrums eines zu identifizierenden Mikroorganismus. Vorteilhafterweise wird dieses Probenspektrum SAM unter möglichst identischen Messbedingungen, insbesondere gleicher Matrix, wie die Referenzspektren  $REF$  in Schritt S1 gemessen. Nachfolgend (Schritt S7) erfolgt eine Ähnlichkeitsanalyse des Probenspektrums SAM mit den in der Datenbank DB enthaltenen synthetischen Referenzspektren  $REF_s$ . Auch dieser Spektrenabgleich kann visuell, vorzugsweise jedoch computergestützt, erfolgen. Dabei kann prinzipiell auf die gleichen oder ähnliche Algorithmen zurückgegriffen werden, die bereits in Schritt S3 zur Ermittlung der empirischen Signale  $S_{em}$  Anwendung fanden. Durch die Einfachheit der verwendeten Referenzspektren  $REF_s$  kann der Spektrenabgleich in Schritt S7 auch mit sehr einfachen Algorithmen durchgeführt werden, die sich lediglich auf einen Vergleich der Signale  $S_{id}$ ,  $S_{em}$  der synthetischen Referenzspektren  $REF_s$  beschränken. Selbstverständlich müssen hier Toleranzkriterien vorgegeben werden, die festlegen, um wie viel ein in Frage kommendes Signal  $S_{id}$ ,  $S_{em}$  des Probenspektrums SAM hinsichtlich der Masse und/oder der Intensität abweichen darf, um als übereinstimmend gewertet zu werden. Ferner sollte eine Mindest-Anzahl übereinstimmender Signale für eine Identifizierung vorgegeben werden. Im Schritt S8 erfolgt eine Ausgabe eines Resultats. Dieses kann entweder darin bestehen, dass eine Übereinstimmung des Probenspektrums SAM mit einem der Referenzspektren  $REF_s$  erkannt wurde. Ferner kann auch eine Information über den Grad der Übereinstimmung ausgegeben werden. Handelt es sich – wie häufig der Fall – bei der Probe des unbekannten Organismus um eine Mischkultur, so ermöglicht das Verfahren sogar die Identifizierung mehrerer, nebeneinander vorliegender Mikroorganismen. In einem solchen Fall werden als Ergebnis mehrere Treffer ausgegeben. Stimmt das Probenspektrum SAM mit keinem der synthetischen Referenzspektren  $REF_s$  innerhalb der zugelassenen Toleranzen überein, das heißt, der unbekannte Mikroorganismus ist nicht in der Datenbank DB enthalten, so ist auch dieses ein ausgebbares Resultat.

[0025] In bestimmten Fällen, in denen ein Mikroorganismus nicht oder nicht eindeutig identifiziert werden kann, kann vorteilhaft vorgesehen sein, dass im Anschluss an den Schritt S7 ein zusätzlicher Direktvergleich mit den gemessenen "natürlichen" Referenzspektren  $REF$  durchgeführt wird. Dieses Vorgehen entspricht der bekannten Vorgehensweise zur Identifizierung von Mikroorganismen mittels MALDI-TOF-MS.

[0026] Durch die Einfachheit und den konzentrierten Informationsgehalt der synthetischen Referenzspektren  $REF_s$  zeichnet sich das erfindungsgemäße Verfahren durch eine wesentlich höhere Zuverlässigkeit aus.

#### Bezugszeichenliste

S1 Messung von Referenzspektren  
S2 Signalidentifizierung  
S3 Ermittlung empirischer Signale

S4 Erzeugung eines synthetischen Referenzspektrums  
S5 Datenbankerstellung  
S6 Erfassung eines Probenspektrums  
S7 Spektrenabgleich  
S8 Ergebnisausgabe  
 $S_{id}$  identifiziertes Signal  
 $S_{em}$  empirisches Signal  
 $REF$  gemessenes Referenzspektrum  
 $REF_s$  synthetisches Referenzspektrum  
DB Datenbank  
SAM Probenspektrum  
RES Ergebnis

#### Patentansprüche

1. Verfahren zur Identifizierung von Mikroorganismen mittels Matrixassisted-Laser-Desorption-Ionisation-Time-of-Flight Massenspektrometrie (MALDI-TOF-MS), wobei eine Datenbank mit Referenzspektren, bestehend aus Massenspektren einer Anzahl bekannter Mikroorganismen, erstellt wird, ein MALDI-TOF-Massenspektrum (Probenspektrum) einer Probe eines zu identifizierenden Mikroorganismus aufgenommen wird, eine Ähnlichkeitsanalyse des Massenspektrums des zu identifizierenden Mikroorganismus mit den in der Datenbank enthaltenen Referenzspektren durchgeführt wird,

**dadurch gekennzeichnet**, dass als Referenzspektren ( $REF$ ) synthetische Massenspektren ( $REF_s$ ) verwendet werden, die durch Zusammenfassung einer gegenüber natürlichen Massenspektren reduzierten Anzahl von für den jeweiligen Mikroorganismus spezifischen Signalen ( $S$ ) erzeugt werden.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass die Signale ( $S$ ) eines synthetischen Referenzspektrums ( $REF_s$ ) mindestens ein identifiziertes Signal ( $S_{id}$ ) umfassen, welches einem einzelnen molekularen Zellbestandteil des jeweiligen Mikroorganismus zugeordnet wurde.

3. Verfahren nach Anspruch 2, dadurch gekennzeichnet, dass der Zellbestandteil ein Peptid, ein Protein, eine Ribonukleinsäure und/oder ein Lipid ist.

4. Verfahren nach einem der Ansprüche 2 oder 3, dadurch gekennzeichnet, dass der Zellbestandteil ein ribosomales Protein, insbesondere ein Protein der großen Ribosomenuntereinheit ist.

5. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass die Signale ( $S$ ) eines synthetischen Referenzspektrums ( $REF_s$ ) mindestens ein empirisches Signal ( $S_{em}$ ) umfassen, welches durch Vergleich von Massenspektren bekannter Mikroorganismen als spezifisch für einen Mikroorganismus ermittelt wurde.

6. Verfahren nach Anspruch 5, dadurch gekennzeichnet, dass der Vergleich visuell und/oder computergestützt durchgeführt wird.

7. Verfahren nach einem der Ansprüche 5 oder 6, dadurch gekennzeichnet, dass als Kriterien zur Ermittlung eines empirischen Signals ( $S_{em}$ ) eine vorgebbare Mindest-Häufigkeit für ein Auftreten des Signals in einer Anzahl von Massenspektren des selben Mikroorganismus und eine vorgebbare Mindest-Intensität des Signals vorgegeben werden.

8. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass ein Signal ( $S$ ) durch ein Koordinatenpaar, bestehend aus einer Masse ( $m$ )

oder einem Masse-Ladungsverhältnis ( $m/z$ ) als x-Koordinate und einer absoluten oder relativen Intensität als y-Koordinate, repräsentiert wird.

9. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass die Anzahl der Signale (S) eines synthetischen Referenzspektrums (REF<sub>S</sub>) 5 bis 30, insbesondere 10 bis 15, beträgt.

10. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass die Aufnahme der Massenspektren von nicht vorbehandelten Zellen durchgeführt wird.

11. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass die Ähnlichkeitsanalyse des Massenspektrums des zu identifizierenden Mikroorganismus mit den in der Datenbank (DB) enthaltenen Referenzspektren (REF<sub>S</sub>) sich auf einen Vergleich der in den Referenzspektren (REF) enthaltenen Signale (S) beschränkt.

12. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass für die Ähnlichkeitsanalyse den in den Referenzspektren (REF<sub>S</sub>) enthaltenen Signalen (S) Gewichtungen zugeordnet werden.

13. Datenbank, erhältlich durch ein Verfahren nach einem der Ansprüche 1 bis 12.

14. Verwendung einer Datenbank nach Anspruch 13 für die Identifizierung von Mikroorganismen mittels MALDI-TOF-MS.

---

Hierzu 6 Seite(n) Zeichnungen

---

30

35

40

45

50

55

60

65

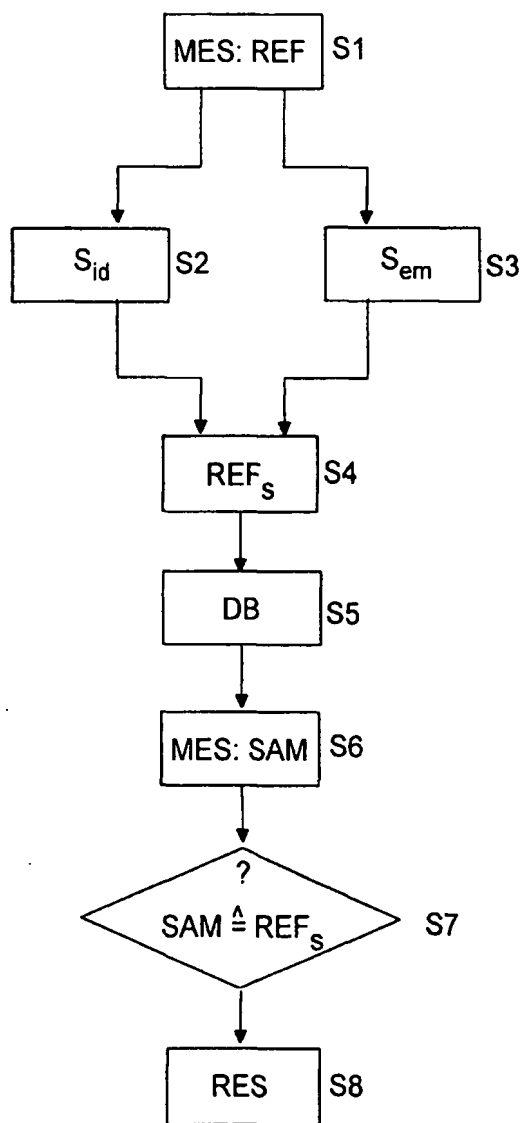


Fig. 1

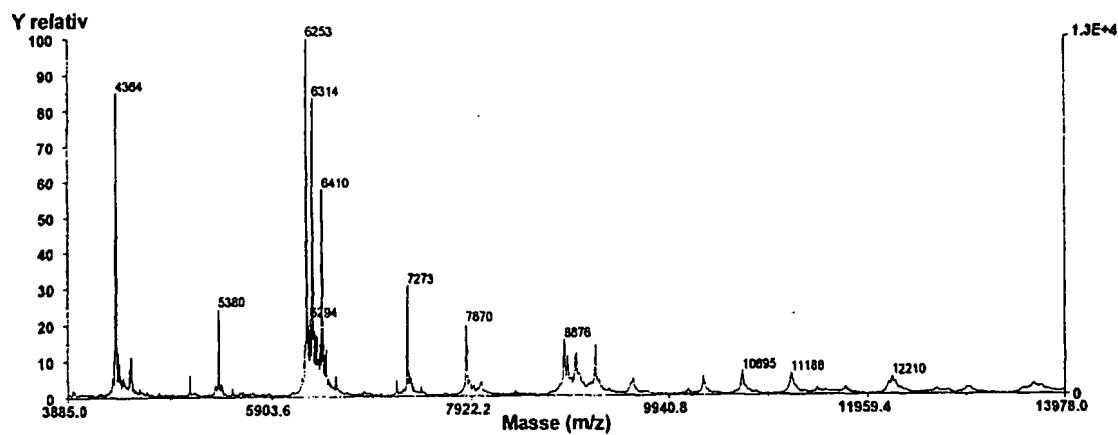
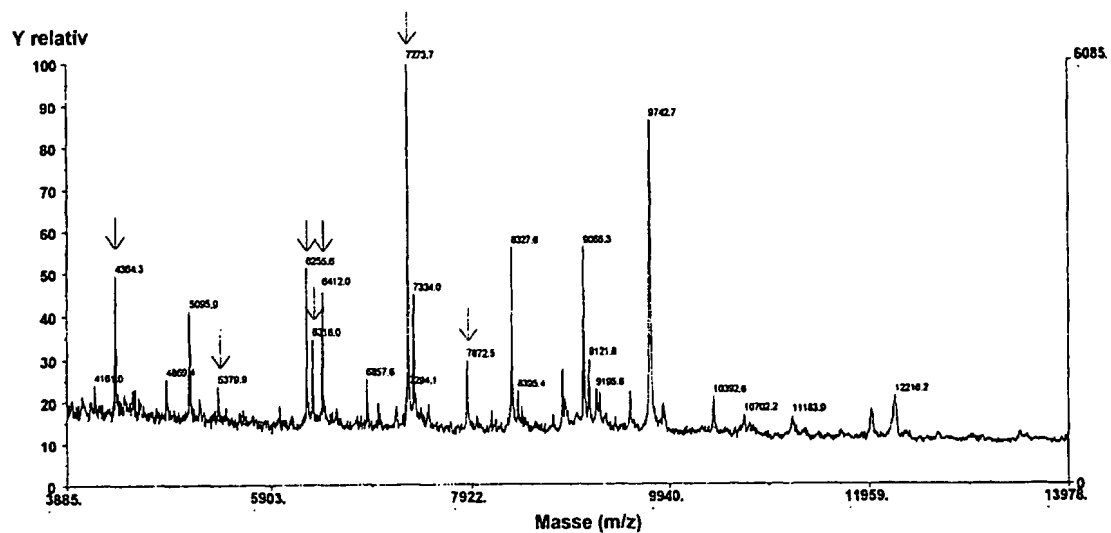
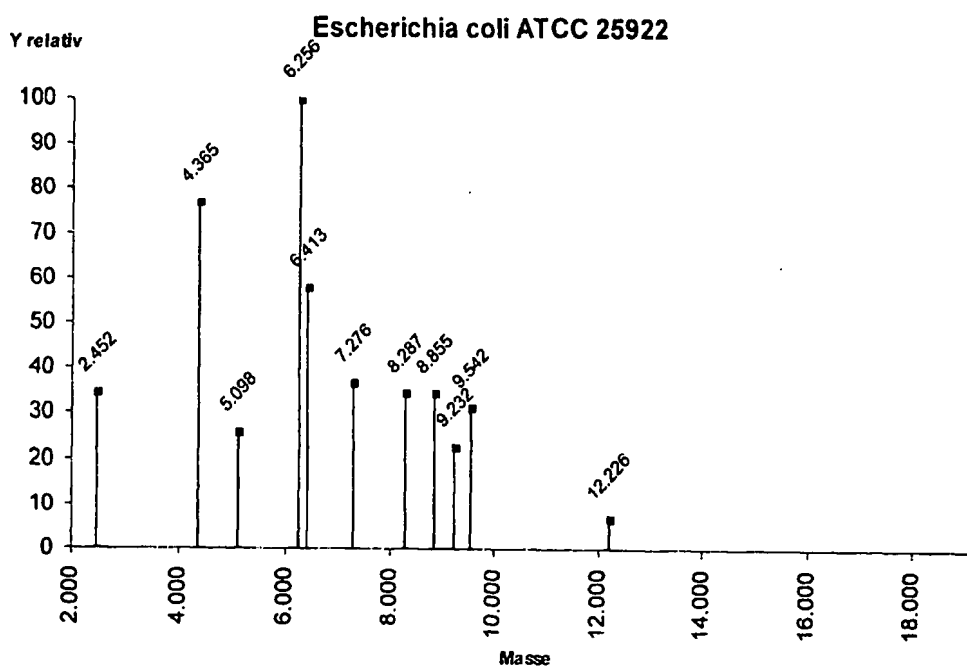


Fig. 2



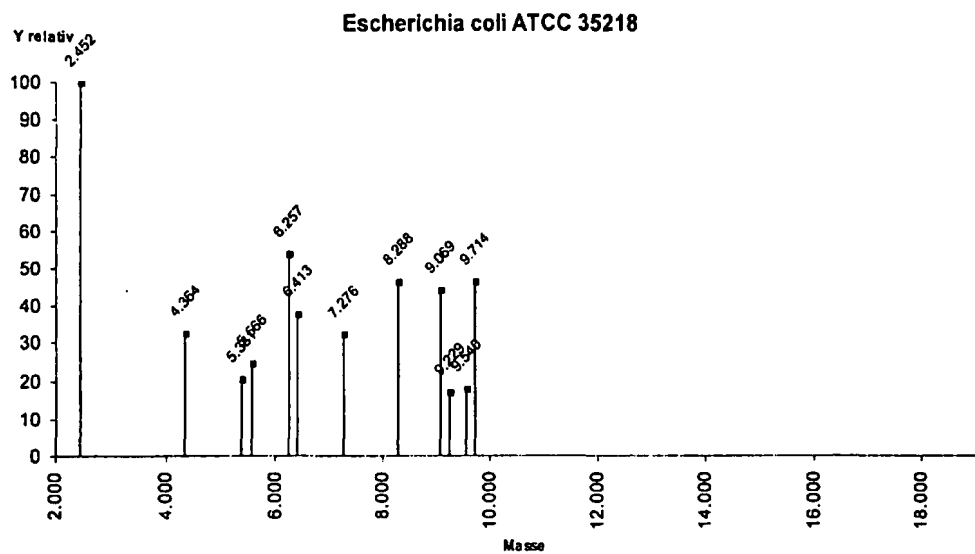
*Escherichia coli* ATCC 25922 $Y_{\max} = 4314$ 

Peak	X	$Y_{\text{absolut}}$	$Y_{\text{relativ}}$
1	2.452	1500	34,77
2	4.365 x	3316	76,87
3	5.098	1120	25,96
4	6.256	4314	100,00
5	6.413 x	2505	58,07
6	7.276 x	1591	36,88
7	8.287	1500	34,77
8	8.855	1500	34,77
9	9.232	983	22,79
10	9.542	1350	31,29
11	12.226	308	7,14

**Fig. 3**

*Escherichia coli* ATCC 35218 $Y_{\max} = 3232$ 

Peak	X	$Y_{\text{absolut}}$	$Y_{\text{relativ}}$
1	2.452	3232	100,00
2	4.364 x	1056	32,67
3	5.381	658	20,36
4	5.566	801	24,78
5	6.257	1743	53,93
6	6.413 x	1227	37,96
7	7.276 x	1046	32,36
8	8.288	1500	46,41
9	9.069	1431	44,28
10	9.229	547	16,92
11	9.540	582	18,01
12	9.714	1500	46,41

Fig. 4

*Klebsiella oxycola* TYP 1 $Y_{\max} = 7600$ 

Peak	X	Y <sub>absolut</sub>	Y <sub>relativ</sub>
1	4.363	6000	78,95
2	5.190	2500	32,89
3	5.408	4000	52,63
4	6.255	6000	78,95
5	6.384	6000	78,95
6	7.330	3000	39,47
7	7.363	3000	39,47
8	7.695	4000	52,63
9	8.265	6000	78,95
10	8.285	7000	92,11
11	9.134	2000	26,32
12	9.154	2000	26,32
13	9.473	1800	23,68

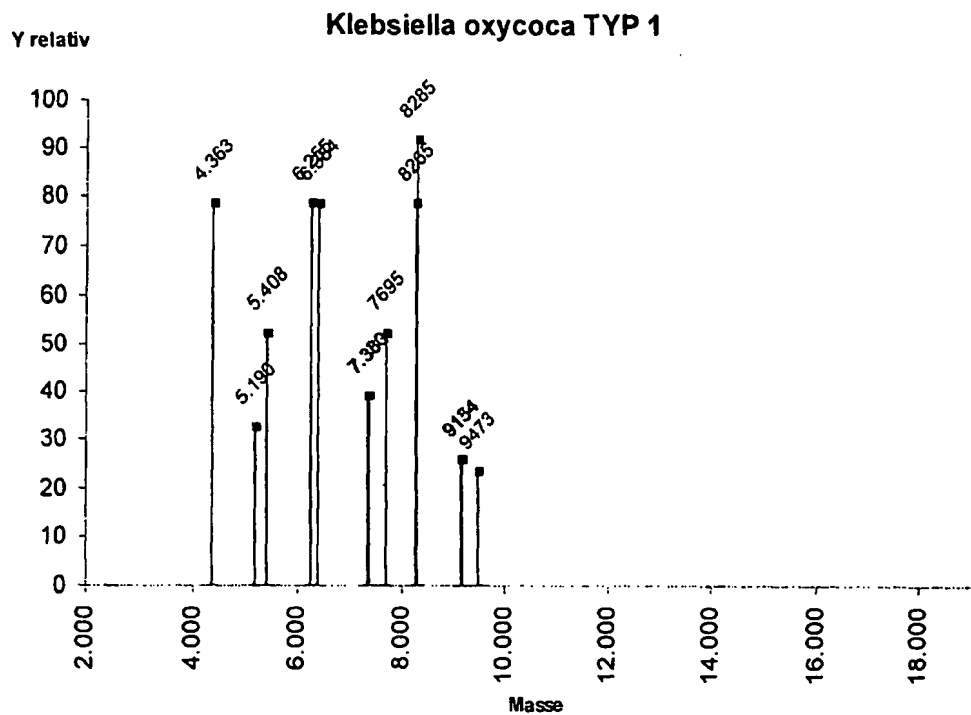
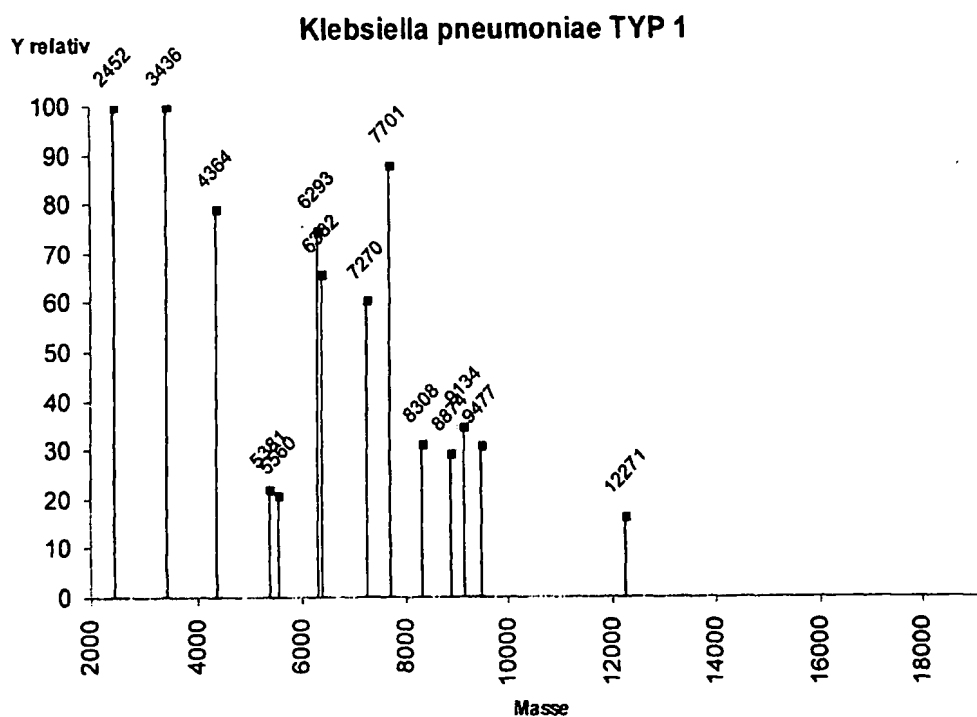


Fig. 5

*Klebsiella pneumoniae* TYP 1 $Y_{\max} = 1500$ 

Peak	X	$Y_{\text{absolut}}$	$Y_{\text{relativ}}$
1	2.452	1500	100,00
2	3.436	1500	100,00
3	4.364	1189	79,27
4	5.381	334	22,27
5	5.560	311	20,73
6	6.293	1121	74,73
7	6.382	986	65,73
8	7.270	905	60,33
9	7.701	1319	87,93
10	8.308	472	31,47
11	8.874	442	29,47
12	9.134	523	34,87
13	9.477	464	30,93
14	12.271	246	16,40

**Fig. 6**

# **Machine translation of DE10038694A**

## **DE10038694A - Title / Abstract**

**Identification of micro-organisms, using MALDI-TOF-MS, uses a synthetic reference spectrum for comparison with the mass spectrum to register the micro-organism under study**

Identifying micro-organisms uses matrix assisted laser desorption ionization/time of flight/mass spectrometry (MALDI-TOF-MS), is new.

A database (DB) of reference spectra has mass spectra of some known micro-organisms.

The reference spectrum (REF) uses synthetic reference spectra (REFS), which gather a reduced number of specific signals for a given micro-organism compared with natural mass spectra.

Identifying micro-organisms by MALDI-TOF-MS, the signals from a synthetic reference spectrum contain at least one identification signal (Sid) relating to a single molecular cell component of each micro-organism.

The molecular cell component is a peptide, protein

, a ribonucleic acid (RNA) and/or a lipid, a ribosomal protein and especially a protein of a large ribosome sub-unit.

The signals from a synthetic reference spectrum contain at least one empirical signal (Sem), specific for a given micro-organism, through comparison of the mass spectra of known micro-organisms.

The comparison is visual and/or computer assisted.

The criteria for the generation of an empirical signal are a given minimum frequency for the presence of a signal in a number of mass spectra for the same micro-organism, and a given minimum signal intensity.

A signal is represented by a pair of coordinates composed of a mass or a mass/charge ratio as the X-coordinate, and an absolute or relative intensity as the Y-coordinate.

The synthetic reference spectrum gives 5-30 signals and preferably 10-15.

The mass spectra are registered from cells which have not been prepared.

The analysis of the similarity of the mass spectra of the micro-organisms to be identified with the reference spectra stored in the database is restricted to a comparison of the signals contained in the reference spectra.

The analysis of the similarity of the signals in the reference spectra uses weighting.

**Claims:**

1. Procedure for the identification of micro organisms by means of Matrixassisted laser desorption ionization time OF Flight mass spectrometry (MALDI TOF ms), how

a data base with reference spectra, consisting of mass spectra of a number of well-known micro organisms, is provided,

a MALDI TOF mass spectrum (sample spectrum) of a sample one is taken up microorganism which can be identified,

a similarity analysis of the mass spectrum of the microorganism with the reference spectra, contained which can be identified, in the data base is accomplished,

is characterized by that as reference spectra (REF) synthetic mass spectra (REFS) are used, which reduced opposite natural mass spectra by summary number of signals (s) specific for the respective microorganism are produced.

2. Procedure according to claim 1, is characterized by that the signals (s) of a synthetic reference spectrum (REFS) cover at least an identified signal (simmer), which was assigned to an individual molecular cell component of the respective microorganism.

3. Procedure according to claim 2, is characterized by that the cell component is a Peptid, a protein, a Ribonukleinsaeure and/or a Lipid.

4. Procedure after one of the claims 2 or 3, is characterized by that the cell component is a ribosomales protein, in particular a protein of the large Ribosomenuntereinheit.

5. Procedure after one of the preceding claims, is characterized by that the signals (s) of a synthetic reference spectrum (REFS) cover at least an empirical signal (Sem), which micro organisms well-known by comparison of mass spectra as specific for a microorganism were determined.

6. Procedure according to claim 5, is characterized by that the comparison is accomplished visually and/or computer-assisted.

7. Procedure after one of the claims 5 or 6, is characterized by that as criteria for the determination of an empirical signal (Sem) a minimum frequency given in advance for an occurrence of the signal in a number of mass spectra of the same microorganism and a minimum intensity given in advance of the signal are given.

8. Procedure after one of the preceding claims, is characterized by that a signal (s) is represented by a pair of coordinates, consisting of a mass (m) or a mass charge relationship (m/z) as x-coordinate and an absolute or relative intensity as Y-coordinate.
9. Procedure after one of the preceding claims, is characterized by that the number of signals (s) of a synthetic reference spectrum (REFS) amounts to 5 to 30, in particular 10 to 15.
10. Procedure after one of the preceding claims, is characterized by that the admission of the mass spectra of not pre-treated cells is accomplished.
11. Procedure after one of the preceding claims, is characterized by that the similarity analysis of the mass spectrum of the microorganism with the reference spectra (REFS), contained which can be identified, in the data base (railways), limited to a comparison of the signals (s) contained in the reference spectra (REF).
12. Procedure after one of the preceding claims, is characterized by that for the similarity analysis weightings are assigned to the signals (s) contained in the reference spectra (REFS).
13. Data base, available by a procedure after one of the claims 1 to 12.
14. Use of a data base after claim 13 for the identification of micro organisms by means of MALDI TOF ms.

## Description:

The invention concerns a procedure for the identification of micro organisms by means of matrix assisted laser desorption ionization Time OFFlight mass spectrometry (MALDI TOF ms) with the characteristics specified in the generic term of the claim 1 as well as in the procedure provided and for the procedure usable data base.

A fast and reliable identification from micro organisms is in different ranges of the health service, for example the diagnostics of infections, as well as the foodstuffs industry of crucial importance. The traditional identification by means of the direct bacteria proof requires first the Anzuechtung of the micro organisms from a sample of a material, which can be identified. Following microscopic investigation procedures serve mainly provisional orientation over the bacteria content as well as the micro morphology and/or coloring characteristics of the clinical investigation sample. The identification from Isolaten to the species level requires pretty often a Subkultivierung for the production of a pure culture. In the classical identification procedure of the medical microbiology using a suitable combination of differential media (so-called "multicolored row"), specific metabolic achievements of the microorganism which can be identified are finally seized. Hauptnachteil of the mikrobiellen procedure is its very high time requirement. More modern molecular-biological beginnings, like the PCR method (polymerase nuclear chain reaction) and the 16S-rRNA method, involve the genetic analysis of the gene COM and/or certain Ribonukleinsaeuren isolated before. These procedures gained because of their high sensitivity strongly significance. They require a cultivation of the organisms in the laboratory just like the micro-biological characterisation and are beyond that loaded with a substantial personnel and instrumental expenditure. Furthermore a infrared-spectroscopic procedure was developed, with that vibration spectra of intact cells ("finger print spectra") in FTIR spectrometers is noted and with a data base with vibration spectra of well-known micro

organisms adjusted. This still very new technology still is in the development and can be accomplished at present only by particularly trained and experienced personnel, so that this beginning could not become generally accepted yet in practice.

With the so-called MALDI TOF Massenspektrometrie in the last years a procedure was developed, which is accessible to the analysis of biological macromolecules contrary to conventional measure-spectrometric procedures also. With the MALDI TOF Ms technology the sample with a usually crystalline organic compound, to the so-called matrix, which can be examined, is given on a Probenplatte, whereby the sample is built into the matrix crystals, and brought with a laser beam in reciprocal effect. Individual molecules of the sample are desorbed and ionized by the sample carrier. Subsequently, in such a way produced ions in an electrical field are accelerated and its flying time (time OF Flight) up to reaching a detector is registered. Since acceleration depends on the mass of an ionized molecule, the flying times reflect the molecule masses existing in the sample. The MALDI TOF Ms technology finds nowadays mainly in the area of protein analytics ("Proteomics") and with the RNA and DNA analytics application. It was suggested using the method for the identification of micro organisms into analogy to the infrared-spectroscopic finger print method described above. But the MALDI TOF mass spectrum of a cell excerpt or intact cells (for example WHERE 98/09314) of the unknown microorganism is compared with the spectra of well-known organisms. The comparison of the sample spectrum with the reference spectra of the data base takes place usually computer-assisted by means of statistic-mathematical algorithms, which were developed in so-called pattern recognition procedures. The mass spectra resemble each other increasing with increasing relational proximity of the micro organisms. That is, the probability of a repetition of certain signals in the mass spectra increases. Therefore the procedure is with superordinate classification levels, approximately at kinds or families, with a increased uncertainty connected, if an allocation on master level fails, for example because no reference spectrum of the trunk in the data base is present. The procedure is dependent therefore on very extensive data bases with a multiplicity of representative reference spectra of well-known trunks. In addition the difficulty comes to discriminate against a spectral background noise from genuine signals to.

Task of the present invention is it therefore to develop further would genericin accordance with-eat MALDI TOF Ms procedures for the identification of micro organisms going by that its reliability is increased, without the necessary process expenditure to increase. It is furthermore in the procedure provided and for the procedure applicable data base to be made available.

This task is solved by a procedure with the characteristics of the claim 1 as well as by a data base and its use after the Claims 13 and 14.

Procedure in accordance with it is intended that as reference spectra synthetic mass spectra are used, which reduced natural" (did not reduce) mass spectra by summary of one opposite "number of signals specific for the respective microorganism are produced. By the reduction of the reference spectra on a relatively small number of characteristic signals, a substantial data and information reduction are reached. Not only storage location is saved by the data reduction, but also a temporal expenditure for data communication, so that the procedure also in principle for an application over interlaced data-processing systems (e.g. Internet) is suitable. Beyond that the information reduction of the reference spectra makes a clear acceleration of the similarity analysis of the mass spectrum of the microorganism with the reference spectra, which can be identified, for the data base possible, since the analysis can be limited now to a comparison of the signals contained in the reference spectra.



Compared to conventional procedures, in which a sample spectrum of the organism with "natural" reference spectra, which can be identified, is adjusted, is a reliability of the procedure according to invention, i.e. the probability, a correct allocation of the unknown organism to meet, clearly increases. Herein the bedeutendste advantage of the invention is to be seen. The increased reliability is to be explained by the high concentration of specific information in the reference spectra, which is not covered by signals of small significance or noise. Also with a possible failure with an allocation on master level of a sample, about because no reference spectrum of this trunk is available in the data base, the procedure supplies reliable identification on superordinate classification levels, for example on generic or kind level. The sensitivity of the proceeding according to invention is also not impaired by differences in different spectra, which are caused for example by different cultivation conditions, different cell stages or deviating signal of intoxication conditions.

A further advantage of the procedure represents the small necessary sample quantity of the microorganism which can be identified, which is available in shorter cultivation times. Furthermore also Mischkulturen can be analyzed, so that without the rearing of pure cultures can be done.

The signals contained in the synthetic reference spectra can be differentiated with respect to two categories. A particularly favourable arrangement of the invention plans that the signals of a reference spectrum cover at least an identified signal, which was clearly assigned to a characterized molecular cell component of the respective microorganism. For the identification of a microorganism suitable cell components are for example certain Peptide, proteins, Ribonukleinsäuren and/or Lipide. As particularly favourable, a signal allocation of a ribosomalen protein, in particular a protein of the large Ribosomenuntereinheit proved to make. For bacteria this proteins of the so-called 50S-Untereinheit are and with mushrooms the 60S subunit. Proteins represent a main part of mikrobieller cells. This applies in particular for Ribosomenproteine, which are independently of a development stage of the cell, a nutrient offer or other cultivation conditions constantly present and thus reliable signals in the mass spectra represent. In addition it comes that amino acid sequences of similar proteins of different species or even different trunks differ at least slightly from each other. Therefore the similar proteins exhibit different masses and are suitable for their distinction. Thus the allocation of three mass signals in the mass spectrum of Escherichia succeeded to the inventors for example for the first time coli ( $m/z = 4365, 6413, 7276$ ) at the proteins L29, L30 and L36 of the large 50S-Ribosomenuntereinheit. These signals are constant components in mass spectra of E. coli, not however from most other micro organisms. They are suitable therefore particularly for the identification by Escherichia coli and for the admission into a synthetic reference spectrum for E. coli.

In further training of the present invention it is intended that the signals of a synthetic reference spectrum cover an empirical signal, which micro organisms well-known by comparison of a multiplicity of mass spectra as specific for a microorganism was determined as the second signal category at least. It concerns signals, whose origin, i.e. their causing molecular cell component do not admit, are, which are rated however due to certain criteria as characteristic for a microorganism. Preferably these criteria cover a minimum frequency given in advance for the occurrence of a signal in a number of mass spectra of the same microorganism as well as an average minimum intensity given in advance of the signal. The minimum frequency should amount to at least 50%, in particular at least 70%, preferably at least 90%. The determination of the empirical signals by comparison of measured mass spectra can be accomplished visually, preferably however computer-assisted. Appropriate algorithms (pattern recognition procedures) are well-known and are not more near to be described here. The number of signals of a reference spectrum amounts to favourable-proves 5 to 30. In practical attempts in particular a number from 10 to 15 proved as sufficient. Furthermore it is preferentially intended that a signal of a synthetic reference spectrum is represented by only a pair of coordinates. The pair of coordinates consists mass charge of a mass and/or a

relationship as x-coordinate on the one hand and an absolute or relative intensity as Y-coordinate on the other hand. Compared to "natural" mass spectra, which contain about 16000 data points, this means a substantial data reduction. For the alignment of a sample spectrum with the reference spectra of the data base favourable-prove to the individual identified and empirical signals of the synthetic reference spectra weightings can according to their significance to be assigned.

Furthermore the invention covers a data base, which were provided with the procedure according to invention, as well as the use of the data base for the identification of micro organisms by means of MALDI TOF ms.

The invention is more near described in the following in remark examples on the basis the associated designs. Show:

Fig. 1 a flow chart of a typical execution of the procedure according to invention;

Fig. 2 MALDI TOF mass spectra of Escherichia coli and of isolated Ribosomen from Escherichia coli and

Fig. 3 to 6 synthetic reference spectra of two Escherichia coli trunks as well as two Klebsiella species.

Fig. 1 shows a typical operational sequence according to invention in a flow chart. In a first step S1 a if possible large number of reference spectra REF of well-known micro organisms is measured. In the following some details are given for sample preparation and for data acquisition.

#### Sample preparation and MALDI TOF Datenakquisition

For an MALDI TOF analysis 5 to 100  $\mu$ g wet cells or also 5 to 50  $\mu$ g dried cells of a well-known or bacterium or mushroom which can be identified are zirka needed. A pretreatment of the cells, for example a cell explanation, is not necessary. The wet cells can be transferred either directly by a agar culture with a sterile inoculation eye to one also Template Probenplatte specified. Alternatively cells of a liquid culture also abzentrifugierte can be used. Subsequently, the cells on the Probenplatte are mixed with 0,2 to 1  $\mu$ l of a matrix solution. Into modification of this procedure the wet cells can be mixed also before their transmission to the Probenplatte with the matrix solution and transferred as suspension. For the following measurements a matrix solution from 100 mg/ml 2,5-Dihydroxybenzoesaeure was used in one mixture from 50% acetonitrile and 50% water with 3% tri fluorine acetic acid. Other well-known matrix solutions are likewise suitable. After drying process of the sample, which accompanies with a crystallization, the sample can be subjected directly to a MALDI TOF massenspektrometischen analysis. For the available measurements each point of sample of a Probenplatte in a conventional mass spectrometer with zirka 50 to 300 laser pulses of a nitrogen laser with a wavelength of 337 Nm became lively. The acquisition of positive ion mass spectra took place in the linear measuring mode in a mass range from 2000 to 20000 m/z. A typical example of a in this way received positive MALDI TOF of mass spectrum REF von Escherichia coli is in the top the Fig. 2 within the mass range from zirka m/z = 4000 to 14000 represented.

In the following step S2 (Fig. 1) is accomplished a peak allocation to certain molecular cell components. A possible proceeding, which falls back to well-known methods of biochemistry and molecular biology, is described briefly in the following by the example by Ribosomenproteine of the large Ribosomenuntereinheit.

Identification of unknown quantities of signals

A protein excerpt of a cell culture is isolated by means of a 2D-Gelelektrophorese. Submitted in question coming so-called Proteinspots afterwards a tryptischen digest, with which the protein is split enzymatically into small protein fragments. If anti-bodies are present against Ribosomen, the relevant Proteinspots can be recognized also by a Immunoassay (for example Western Blot analysis). By the tryptischen digest won protein fragments by means of a HPLC, which is suitable for particularly small sample quantities, are then isolated and ansequenziert. With in this way determined sequence fragments can then be tried to identify the Ribosomengene in a data base. If appropriate genes are present in the data base, the corresponding protein mass can be determined from the entire gene sequence. If an appropriate gene in the data base cannot be found, then the entire gene with well-known means of molecular biology, which are not to be implemented here more near, which for the appropriate ribosomale protein are coded, isolated and sequenziert, must. If the gene sequence is well-known, the translation in the protein sequence and determination of the theoretical protein mass follows. An examination of this theoretical mass can take place, as the appropriate protein poet of the 2D-Gelelektrophorese of a MALDI TOF Massenspektrometrie is subjected. In the case of deviations from the theoretical protein mass modifications of the protein can be determined by MALDI TOF analysis of the tryptischen Verdaus.

For elucidation is in the Fig. 2 in the lower part a mass spectrum of the 70S-Ribosomens from Escherichia coli, which was isolated by means of 2D-Gelelektrophorese, represented. With 70S the entire Ribosomen of a protein is marked, which consists of the large 50S-Untereinheit and the small 30S-Untereinheit. Both subunits consist for their part of a set of proteins, which are named the letter L (for large) and the letter S (for small). It is easily evident that a number of the most intensive signals from the spectrum of Escherichia coli (see Fig. 2 top) to ribosomale proteins to attribute leaves themselves. Appropriate signals are marked in the illustration by arrows. The inventors succeeded in for the first time assigning three of these signals to certain proteins of the large 50S-Ribosomenuntereinheit. It concerns the proteins L29, L30 and L36 with the masses  $m/z = 7274, 6412$  and  $4364$ . Since these signals with a high reliability in the mass spectra of Escherichia emerge coli, they are particularly suitable for the identification. It used as identified signals simmer for the synthetic reference spectrum of REFS for Escherichia coli.

In an alternative or additional step S3 (Fig. 1) empirical signals  $S_{em}$  from the measured reference spectra of REF of the well-known micro organisms are determined. The determination of the empirical signals  $S_{em}$  takes place via comparison of a multiplicity of mass spectra, which were noted by the same trunk, among themselves as well as via comparison of these mass spectra with those from other organisms. Such signals are determined as characteristic for a microorganism, those if possible frequently in the mass spectra of the same organism to emerge and if possible rarely in the mass spectra of another. For the frequency of an occurrence of a suitable empirical signal  $S_{em}$  thereby a minimum value, for example  $>$  related to all spectra of an organism, can be given to 70%. Likewise such a signal should possess a minimum intensity given in advance, in order to facilitate its distinction from the background noise to. The determination of the empirical signals  $S_{em}$  in the step S3 can take place via visual spectrum comparison via the user. It is however preferentially intended accomplishing this step automated with the help of suitable computer programs. Here for instance an algorithm is applicable, which examines the measured reference spectra of REF regarding the criteria mentioned. There is however also, for example well-known from the infrared spectrometry, computer-assisted procedures, which are with application of statistic algorithms able to recognize and filter recurring signals.

In the steps the S2 and S3 determined identified and empirical signals simmer,  $S_{em}$  in a following step S4 into synthetic reference spectra of REFS are combined. For each well-known microorganism a reference spectrum of REFS is produced. Into the Fig. 3 to 6 is represented exemplarily such synthetic reference spectra of REFS for

two *Escherichia coli* trunks (ATCC 25922 and ATCC 35218) and two *Klebsiella* species - *Klebsiella pneumoniae* and *Klebsiella oxytoca* -. In each case everyone is the Fig in the top. 3 to 6 the synthetic reference spectrum of REFS in coordinate form represented, while in the lower part the graphic representation is shown by mass spectra in the typical form in each case. The synthetic reference spectra of REFS cover in the examples shown eleven to fourteen signals  $S$ . thereby are in the *Escherichia coli* Referenzspektren REFS the signals identified as the proteins L29, L30 and L36 of the large 50S-Ribosomenuntereinheit von *Escherichia coli* simmer marked. These signals simmer emerge in both trunks of *Escherichia coli*. They are suitable for the identification of the kind and for the demarcation in relation to *Klebsiella* species, in which only one part of the identified signals simmer occurs. In particular becomes with *Klebsiella* (Fig. 5 and 6) the signal with the mass  $m/z = 6413$  does not observe. The distinction of the two *Escherichia coli* trunks (Fig. 3 and 4) by such empirical signals  $S_{em}$  is made possible, which is observed only with one of both. A comparison in the Fig. it shows 3 and 4 synthetic reference spectra shown of REFS that the procedure according to invention is sensitive enough, in order to make a distinction possible of micro organisms even on master level. Besides also a distinction between *Escherichia* is made possible *coli* on the one hand and *Klebsiella* species on the other hand.

In the step the S4 (Fig. 1) produced synthetic reference spectra of REFS are combined in a following step S5 into a data base railways. Within the data base railways the synthetic reference spectra of REFS can be arranged in logical way, for example according to family, kind, kind and trunk. The operational sequence from S1 to S5, i.e. the production of the data base railways, must be accomplished only in principle once. To this data base railways can be fallen back then with the following steps again and again. Of course the data base railways should be constantly extended and updated by reference spectra of REFS of further micro organisms. Like that is about conceivable, reference spectra of REFS von Mutanten, in particular such with changed resistance and/or Virulenzverhalten to take up if these trunks supply characteristic signals. Also the existing reference spectra of REFS should be constantly optimized, for example by the admission of new signals simmer,  $S_{em}$  and/or by deletion of signals on basis of statistic evaluation of the analyzed samples.

In the step S6 then the acquisition of a mass spectrum one takes place microorganism which can be identified. Favourable way is measured this sample spectrum of SAM under if possible identical Messbedingungen, in particular same matrix, as the reference spectra of REF in step S1. In the following (step S7) a similarity analysis of the sample spectrum of SAM with the synthetic reference spectra of REFS contained in the data base railways takes place. Also this spectrum alignment can take place visually, preferably however computer-assisted. Can resemble in principle to or similar algorithms fall back, which already found in step S3 for the determination of the empirical signals  $S_{em}$  application. The spectrum alignment can be accomplished by the simplicity of the used reference spectra of REFS in step S7 also with very simple algorithms, which only on a comparison of the signals simmer yourself,  $S_{em}$  of the synthetic reference spectra of REFS to limit. Of course tolerance criteria must be given, which specify, in order like much a which is applicable signal simmer, to  $S_{em}$  of the sample spectrum of SAM here regarding the mass and/or the intensity to deviate may, in order as agreeing to be rated. Furthermore a minimum number of agreeing signals should be given for an identification. In the step S8 an expenditure of a result takes place. This can consist either of the fact that an agreement of the sample spectrum of SAM with one of the reference spectra of REFS was recognized. Furthermore also information can be spent over the degrees of the agreement. If it concerns - like frequently the case - with the sample of the unknown organism a Mischkultur, then the procedure makes possible even the identification of several, next to each other available micro organisms. In such a case as result several hits are spent. If the sample spectrum of SAM agrees with none of the synthetic reference spectra of REFS within the certified tolerances, i.e., the unknown microorganism is contained not in the data base railways, then also this is a result capable of being output.

In certain cases, in which a microorganism cannot be identified not or clearly, it can be favourably intended that following the step S7 an additional direct comparison with the measured "natural" reference spectra of REF is accomplished. This procedure corresponds to the well-known proceeding for the identification of micro organisms by means of MALDI TOF ms.

By the simplicity and the concentrated information content of the synthetic reference spectra the procedure according to invention is characterised REFS by a substantially higher reliability.

Reference symbol list

S1 measurement of reference spectra

S2 signal identification

S3 determination of empirical signals

S4 production of a synthetic reference spectrum

S5 data base production

S6 collection of a sample spectrum

S7 spectrum alignment

S8 expenditure for result

Simmer identified signal

Sem empirical signal

REF measured reference spectrum

REFS synthetic reference spectrum

Railways data base

SAM sample spectrum

RES result